Calibrating a first and a second microphone

The invention relates to a method of calibrating a first microphone and a second microphone, comprising

- an acquisition step in which a first input audio signal is acquired by means of the first microphone and a second input audio signal is acquired by means of the second microphone;
- a calibration step in which a first sensitivity of the first microphone and a second sensitivity of the second microphone is determined.

The invention also relates to an apparatus comprising a first microphone and a second microphone for acquiring a first and a second input audio signal respectively, and a processor for determining a first sensitivity of the first microphone and a second sensitivity of the second microphone.

The invention also relates to a computer program for execution by a processor, comprising program code for calibrating a first microphone and a second microphone, comprising

- an acquisition step in which a first input audio signal is acquired by means of the first microphone and a second input audio signal is acquired by means of the second microphone;
- a calibration step in which a first sensitivity of the first microphone and a second sensitivity of the second microphone is determined.

The invention also relates to a data carrier storing a computer program for execution by a processor, comprising program code for calibrating a first microphone and a second microphone, which method comprises

- an acquisition step in which a first input audio signal is acquired by means of the first microphone and a second input audio signal is acquired by means of the second microphone;
- a calibration step in which a first sensitivity of the first microphone and a second sensitivity of the second microphone is determined.

An apparatus for calibrating microphones is known from WO-A-0201915. The known apparatus has a multitude of microphones and is useful for e.g. teleconferencing. The multitude of microphones enables better capture of the speech of a speaker, which leads to higher intelligibility at the receiver side. Algorithms for exploiting the multitude of microphones require an accurate calibration of the microphones. This can be done in the factory in an anechoic chamber, but this is expensive. The known apparatus performs calibration after purchase, which enables the connection and the calibration of additional microphones on the fly. The disadvantage however is that the sensitivity of the microphones is determined as the relation between a predetermined acoustical input signal applied to the microphones and a measured electrical output signal from the microphones.

It is a first object of the invention to provide a method of the kind described in the opening paragraph, for calibrating at least two microphones, which is versatile in its use.

It is a second object of the invention to provide an apparatus of the kind described in the opening paragraph, which is versatile in its use.

It is a third object of the invention to provide a computer program for execution on a processor comprising program code coding the method according to the invention.

It is a fourth object of the invention to provide a data carrier storing a computer program according to the invention.

The first object is realized in that in the calibration step an algorithm is applied which enables determination of the sensitivities, in the absence of a loudspeaker for the generation of the input audio signals. To determine the sensitivities in the known apparatus, a loudspeaker is required, which emits a prespecified sound, which serves as the acoustical input for the microphones. In the method of the present invention the calibration is performed by using an algorithm which allows the microphones to be calibrated with naturally present sound, such as speech from a person- e.g. the person executing the method- or a sound picked up on the street. This makes the method, and the apparatus applying the method, more employable to practical usage cases, since it avoids carrying around a loudspeaker.

In an embodiment of the method, the first and the second input audio signal are processed by an adaptive beamforming filter, and the sensitivities are determined by performing a calculation with weights of the adaptive beamforming filter. Beamforming is a widely used algorithm for obtaining an increased sensitivity in the direction of a speaker,

and/or a reduced sensitivity in the direction of a source of noise, by making use of the input signals of a number of microphones. In the embodiment, use is made of the fact that the sensitivities of the microphones can be inferred from the coefficients of the filters used by the beamformer.

In a more specific embodiment of the method, the algorithm comprises calculating

$\sqrt{\frac{1}{L}\sum_{k=0}^{L-1}\left|W^{0}(\Omega_k)\right|^{2}}$, in which W^0 is the discrete Fourier transform of the weights of the beamforming filter after adaptation, and the sum ranges over a predetermined number L of frequencies Ω_k. Performing a calculation in the Fourier domain makes the determination of the sensitivities more robust.

The second object is realized in that the processor is able to determine the sensitivities, in the absence of a loudspeaker for generating the input audio signal. Often the microphones are integrated in an apparatus which is able to calibrate itself, such a teleconferencing apparatus.

The third object is realized in that in the calibration step an algorithm is applied which enables determination of the sensitivities, in the absence of a loudspeaker for generating the input audio signals.

The fourth object is realized in that in the calibration step an algorithm is applied which enables determination of the sensitivities, in the absence of a loudspeaker for generating the input audio signals.

These and other aspects of the method, the apparatus, the computer program and the data carrier according to the invention will be apparent from and elucidated with reference to the implementations and embodiments described hereinafter, and with reference to the accompanying drawings, which serve merely as a non limiting illustration.

In the drawings :

Fig. 1 schematically shows a teleconferencing session;

Fig. 2 schematically shows the method of calibrating a first and a second microphone;

Fig. 3 schematically shows a beamforming apparatus;

Fig. 4 schematically shows a microphone calibration apparatus of the prior art;

Fig. 5 schematically shows an apparatus for relative calibration of a first and a second microphone according to the invention; and

Fig. 6 shows a data carrier.

In these Figures elements drawn dashed are optional depending on the desired embodiment.

In Fig. 1, a teleconferencing session is shown. A locally present person 107 is communicating with a remote person 109, who is e.g. shown on a display 111. For the communication of speech, an audio communication device is required, which is represented by the console 101. It can contain e.g. buttons, a small status display, a loudspeaker for the reproduction of speech uttered by the remote person 109, and a microphone. Practice shows that the locally present person 107 has to be very close to the microphone in the console 101, if he wants that the remote person 109 understands what he is saying. It is much more practical if the locally present person 107 can reside anywhere he likes. To achieve this, more than one microphone is used, illustrated with the first microphone 103 and the second microphone 105 in Fig. 1. Techniques have been developed to take advantage of the spatial arrangement of multiple microphones, in order to better capture the speech of a speaker. The beamforming technique is explained by means of Fig. 3. There are numerous applications which benefit from beamforming and more particular from the method for the relative calibration of at least two microphones described in this text. One example is voice control. E.g. a television set could be equipped with a remote control based on keywords. Beamforming helps in decreasing the keyword recognition failure rate. Also portable devices can be equipped with more than one microphone.

In Fig. 2, a first input audio signal u1, coming from a first microphone 205, and a second input audio signal u2, coming from a second microphone 207, are acquired during an acquisition step ACQ. Both input audio signals u1, u2 are used in the calibration step CAL to determine a first sensitivity a1 of the first microphone 205 and a second sensitivity a2 of the second microphone 207.

Fig. 3 shows an apparatus 241 which is able to apply filtered-sum beamforming to the output of a number of microphones, e.g. three. A first sound source is speech from speaker 201. Suppose that the speech contains a single wavelength, and that a speech wavefront 233 is planar and parallel to an imaginary line running through the first microphone 205 and the second microphone 207. Each microphone then picks up the same

sound signal. The first microphone 205 converts the sound into a sampled first electrical audio signal u1, and the same applies to the second microphone 207 and further microphones if present. If the sensitivities of the microphones 205 and 207 are equal, the sampled electrical audio signals u1 and u2 are equal. Suppose further that a second sound source 203 produces e.g. music of a single wavelength, with planar wavefronts which impinge on the microphone array under an angle θ. Then a music wavefront 231 arrives at the first microphone 205 earlier than at the second microphone 207. This implies that the electrical audio signals u1 and u2 are samples at different phases of a sinusoid of a certain spatial wavelength λ_S, which is related to the direction of incidence θ and the wavelength λ of the music of the second sound source 203. A highpass spatial filter can be designed which transmits the speech from speaker 201 with an infinite spatial wavelength λ_S, but blocks the undesired interference from the second sound source 203. A spatial filter consisting of a single multiplication coefficient for each microphone is sufficient for fixed position, single wavelength sound sources.

For broadband sound sources emitting more wavelengths, a temporal filter is placed behind each microphone instead of a single multiplication coefficient. E.g. a first temporal filter 221 filters the electrical signal u1 of the first microphone 205. Successive samples of u1 are delayed by delay elements, like a first delay element 227, and the delayed samples are multiplied by filter coefficients, like a second filter coefficient 228, and added together by adders, like a first adder 229. The number of filter coefficients is dependent on how many samples from a sound signal are desirable and on how many computing resources are available. The outputs of the temporal filters 221 and 223 are summed by the spatial summation 230, to obtain a spatiotemporal filter output z. Spatiotemporal filter 240 can be described mathematically by means of equation [1]:

$$z(n) = \sum_{m=1}^{M}\sum_{l=1}^{N} w_m(l)u_m(n-lT) \qquad [1]$$

In equation [1], describing a filtered sum beamformer, n is a discrete time index, l is an index of a filter coefficient w, T is a time difference between samples, and m is a microphone index corresponding to one of the microphones (205 and 207) and temporal filters (221 and 223).

In order to properly filter out the audio signal of the interfering second sound source 203, the filter coefficients have to get the appropriate values during an adaptation

phase of beamforming. If adaptation applies e.g. an algorithm which maximizes the power of z(n), under the constraint that for all frequencies Ω_k the following condition is satisfied :

$$\sum_{m=1}^{M}\left|W_m(\Omega_k)\right|^2 = C(\Omega_k)^2 \qquad [2],$$

in which $W_m(\Omega_k)$ is the discrete Fourier transform of the filter $w_m(n)$ and C a constant, then the optimal filter coefficients after adaptation satisfy equation [3]:

$$W_m^o(\Omega_k) = \alpha(\Omega_k)C(\Omega_k)\frac{H_m^*(\Omega_k)}{\sqrt{\sum_{m=1}^{M}\left|H_m(\Omega_k)\right|^2}} \qquad [3].$$

In equation [3], $H_m^*(\Omega_k)$ is the complex conjugate of the discrete Fourier transform of the acoustic impulse response for the microphone with index m. E.g. the first acoustical impulse response h1 in Fig. 2 is the acoustic impulse response modeling the sound transfer from the speaker 201 to the first microphone 205. $\alpha(\Omega_k)$ is an all-pass term which is common to all temporal filters 221, 223 and 225.

For a planar wavefront like the music wavefront 231, the acoustic transfer functions are propagation delays, and hence for each frequency k and microphone index m, equation [4] applies:

$$\left|H_m(\Omega_k)\right|^2 = 1 \qquad [4].$$

In a reverberant room this model is too simple. The sound traveling directly from e.g. speaker 201 to e.g. the first microphone 205, can interfere constructively or destructively with e.g. a first reflection of the sound from speaker 201 on a nearby wall. This can imply that e.g. at the position of the first microphone 205 there is hardly any sound power present at frequency Ω_k. It is very unlikely that the interference should occur for all possible frequencies Ω_k at the spatial position of a microphone, e.g. the first microphone 205. Hence equation [5] is highly likely to be valid:

$$\frac{1}{2N}\sum_{k=0}^{2N-1}\left|H_m(\Omega_k)\right|^2 \approx 1 \qquad [5].$$

Using equation [5], which guarantees that the sound from speaker 201 is transferred approximately equally to the microphones 205 and 207, it can be proven that the relative sensitivities a_m of each microphone follow from equation [6]:

$$a_m = \sqrt{\frac{1}{2N}\sum_{k=0}^{2N-1}\left|W_m^o(\Omega_k)\right|^2} \qquad [6].$$

Hence it is possible to introduce correction factors – 211 and 213 in Fig. 2 – which multiply with the electrical audio signals u1 and u2, in order to make the microphones 205 and 207 equally sensitive. These correction factors can be calculated as in equation [7]:

$$b_m = \frac{c}{a_m} \qquad [7],$$

in which c is a constant.

The spatiotemporal filter 240 implementing filtered sum beamformer [3] identifies the acoustic transfer functions between a sound source measured during calibration and the microphones upto an unknown error factor which is common to all microphones. The fact that the error factor is common to all microphones allows calibration of the microphones relative to each other.

If the speaker is a certain distance away from the microphones, e.g. sitting in a chair watching a television with multiple microphones for voice command, the method works well because the acoustic impulse responses h1 and h2 are similar to propagation delays, which means that all microphones receive essentially the same sound as input. If there is a strong reverberation from e.g. a nearby wall to a microphone at certain frequencies, the method may work less well. Pathological frequency regions can be discarded by modifying the algorithm, by using equation [8] in place of equation [6]:

$$d_m = \sqrt{\frac{1}{2N_i}\sum_i \sum_{k=k_i}^{k_{i+1}} \left|W_m^o(\Omega_k)\right|^2} \qquad [8]$$

The sum in equation [8] is taken over a number i of frequency intervals $[k_i, k_{i+1}]$, in which e.g. no spuriously large values of $W_m(\Omega_k)$ occur. If the sum covers enough frequencies Ω_k, d_m is still a reliable measure of the relative sensitivity of the m-th microphone. N_i is the total number of frequencies in all the intervals $[k_i, k_{i+1}]$ together. To increase accuracy, it is advantageous to also drop the lowest and highest frequencies from the summation, since some of the microphones can have a spurious behavior in these frequency regions.

Fig. 4 shows a microphone calibration apparatus of the prior art. An electrical loudspeaker audio signal e is sent from a signal source 304 to loudspeaker 301, in which it is converted to sound 302, which is picked up by microphone 303. Microphone 303 converts the sound to an electrical microphone audio signal s. In the known apparatus it is required that both the loudspeaker audio signal e and the microphone audio signal s, are sent to a processor 305, which is able to determine microphone sensitivity 307 from the two audio

signals. In the current invention no loudspeaker audio signal e is required by the calibration algorithm. The input of a sound source like speaker 201 is sufficient.

Fig. 5 shows an apparatus 401 for relative calibration of a first and a second microphone 403 and 405 according to the invention. A processor 407 has access to a first audio signal from a first microphone 403 and a second audio signal from a second microphone 405. It is possible to run an algorithm according to the invention, as e.g. illustrated with Fig. 3, on the processor 407, which e.g. calibrates the microphones 403 and 405 after a certain amount of time to counteract time varying effects like e.g. component aging or temperature related effects. Another option is that a user pushes e.g. a button 409 and initiates the calibration, e.g. every time he takes the apparatus into a different room which has different acoustic impulse responses.

An interesting option is to calibrate only when the sound coming into the microphones is speech, by adding a speech detector.

Fig. 6 shows a data carrier for storage of a computer program for execution on a processor describing a method according to the invention for calibrating a first and a second microphone.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention and that those skilled in the art are able to design alternatives, without departing from the scope of the claims. Apart from combinations of elements of the invention as combined in the claims, other combinations of the elements within the scope of the invention as perceived by one skilled in the art are covered by the invention. Any combination of elements can be realized in a single dedicated element. Any reference sign between parentheses in the claim is not intended for limiting the claim. The word "comprising" does not exclude the presence of elements or aspects not listed in a claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements.

The invention can be implemented by means of hardware or by means of software running on a computer.